THE APPLICATION OF SOCIOMETRIC AND EVENT-HISTORY MODELING TO BIBLIOMETRIC DATA: THE CASE OF TRANSGENE PLANTS

Koenraad Debackere, Bart Clarysse and
Michael A. Rappa

15 February 1994

Sloan WP # 3656-94

Massachusetts Institute of Technology

THE APPLICATION OF SOCIOMETRIC AND EVENT-HISTORY MODELING TO BIBLIOMETRIC DATA: THE CASE OF TRANSGENE PLANTS

Koenraad Debackere, Bart Clarysse and
Michael A. Rappa

15 February 1994 Sloan WP # 3656-94

Alfred P. Sloan School of Management
Massachusetts Institute of Technology
50 Memorial Drive, E52-538
Cambridge, MA 02139

THE APPLICATION OF SOCIOMETRIC AND EVENT-HISTORY MODELING TO BIBLIOMETRIC DATA: THE CASE OF TRANSGENE PLANTS

KOENRAAD DEBACKERE, BART CLARYSSE AND MICHAEL A. RAPPA*

This paper examines the determinants of researcher contribution-spans. The contribution-span is the number of years spanning a researcher's first and last known publications in a field. As a consequence, it serves as a unique and useful measure of researchers' persistence in a field. Based on co-authorship data, several sociometric indices are created and their impact on researchers' persistence in their efforts to develop a technology is examined. Evidence is provided from 2,876 researchers active in the field of transgene plants over eleven-years. The findings lend support to the proposition that an individual researcher's network position is an important determinant of his or her persistence in the development of a new technology.

INTRODUCTION

New scientific and technological disciplines typically emerge as a result of the efforts of researchers engaged in problem-solving activities that lead to the development of new ideas and techniques that constitute scientific and technological knowledge (Allen, 1966; Constant, 1980; Laudan, 1984; Layton, 1974 & 1977; Root-Bernstein, 1989; Rosenberg, 1982). As a consequence, a possible approach to study the processes of scientific and technological advance consists of understanding the creation and diffusion of knowledge among researchers. Widely accepted models of the growth of knowledge view this process as a cumulative progression of ideas and techniques (Crane, 1972; Nelson and Winter, 1982; Sahal, 1981) with only infrequent major disruptions or discontinuities (Dosi, 1982; Kuhn, 1970; Rosenberg, 1982; Tushman and Anderson, 1986).

This cumulative character of the growth of scientific and technological knowledge is important to understand the behavior of the researchers who produce it. More specific, researchers have to master a core stock of (often 'tacit') knowledge before they can contribute to the grow h of their field (Collins, 1974; Polanyi, 1958). Herbert Simon estimates that it may take anywhere up to ten years to become an expert in an area of work. Thus, researchers have to persist in their efforts to develop a new technology in order to contribute to its development. Unfortunately, undue persistence often leads to diminishing marginal yields in a stock of knowledge (Ziman, 1987).

The danger of undue persistence is that the researcher's knowledge base becomes obsolete. Katz and Allen (1982) have provided extensive evidence on how 'persistence' eventually undermines research performance through the not-invented-here (NIH) syndrome. Along similar lines, Gieryn (1978) has shown that 'productive' researchers switch research agendas regularly in order to avoid obsolescence. Not astonishingly,

---

*Koenraad Debackere (senior researcher N.F.W.O.) and Bart Clayrsse are with the Vlerick School for Management, University of Gent. Michael Rappa is with the MIT Sloan School of Management. An earlier version of this paper was presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics, Berlin, September 11-15, 1993. This study was supported, in part, by the Flemish Action Program for Biotechnology.

many scholars have documented the creation of new knowledge by researchers from outside the system (Ben-David and Collins, 1966; Edge and Mulkay, 1976; Gieryn and Hirsch, 1983). These 'outsiders' offer ideas that may run counter to prevailing practice in a field. If the new ideas attain legitimacy, though, they may replace existing traditions of practice.

What then determines researchers' persistence with a technology? This question obviously has clear implications for the management of R&D (Rappa, Debackere and Garud, 1992). More specifically, in this paper, we want to address the question to what extent researchers' positions in a network of R&D collaborations influence their persistence. To this end, we use the literature as a source of data on the length of a researcher's association with a field of technological inquiry to examine persistence behavior (Rappa and Debackere, 1992a). Persistence is hypothesized to depend upon individual, relational and collective aspects of the knowledge creation process that technological change entails. Before describing the methodological approach developed in this paper, we first provide an overview of the knowledge creation process within a researcher community.

COLLECTIVE DIMENSIONS OF KNOWLEDGE CREATION

It is realistic to view researchers as working in collaborative relationships with one another (Constant, 1980; Hagstrom, 1965; Hughes, 1989; Pelz and Andrews, 1966). As technological progress often depends on a synthesis of different competencies, collaboration between researchers becomes imperative to solve the complex problems that researchers individually are not able to address. The creation of knowledge by researchers engaged in collaborative relationships with peers results in a steady accumulation of knowledge that other researchers can build upon. Thus, the development of a new technology is not only a cumulative problem-solving process, but also a collective endeavor. The collective dimension of knowledge creation most obviously appears in the acceptance of practices and procedures (e.g. the 'search heuristics' described by Nelson and Winter, 1982) that become institutionalized within a community of researchers. The outcome of this process of institutionalization is an increase in legitimacy of the technology being developed. This creates a technological momentum by attracting new researchers to the field, which in turn augments the rapidity with which new technological knowledge is created (Debackere and Rappa, 1993).

In our previous research, we have introduced the notion of the researcher community to integrate the various aspects of the knowledge creation and diffusion process (Rappa and Debackere, 1992b). The researcher community is defined as a group of scientists and engineers who are committed to solve a set of inter-related scientific and technological problems. These researchers may be organizationally dispersed in public and private sector organizations but, and this is a vital characteristic of every researcher community, they communicate with each other. Within the community, knowledge creation is a collective endeavor driven by individuals engaged in collaborative

relationships with others. Consequently, persistence is in some way determined by individual, relational, and collective aspects of the knowledge creation and diffusion process within this community.

Previous research (Rappa, Debackere and Garud, 1992; Rappa and Garud, 1992) has explored the directions in which individual and collective aspects of knowledge creation affect the contribution-spans of researchers in three different fields of technology development: cochlear implants, polypropylene catalyst development and EPDM rubber catalyst development. At the individual level, the researcher's cumulative productivity was a highly significant predictor ($p<0.001$) of his or her contribution span. At the collective level, the size of the researcher community was a highly significant predictor of the duration of one's contribution-span ($p<0.001$). A more detailed analysis including a second-order term for the size of the researcher community further revealed a U-shaped relationship between population size and researcher contribution-spans, which was interpreted by the authors as suggesting the existence of a critical mass in a field:

> 'The data indicate that when the community is small, population size is negatively related to the length of contribution-spans and increasingly so until it reaches a size of about 250 individuals, at which point the slope of the curve turns positive. This result suggests that there may be a point of critical mass for the community, where its size is sufficiently large to become significant and thereby increase contribution-spans.' (Rappa and Garud, 1992: 345).

The research reported in this paper intends to take these analyses one step further by studying the influence of relational aspects of the knowledge creation and diffusion process on researcher contribution-spans. These relational aspects capture the extent to which researchers associated with the development of a technology collaborate with one another. Collaboration represents the joint creation of ideas by individual researchers. Students of research communities (Beaver and Rosen, 1979; Edge and Mulkay, 1976; Small and Greenlee, 1986; Stewart, 1990) often found that collaboration between researchers in a field steadily increases as the field matures. Consequently, Beaver and Rosen (1978) concluded that the growth of collaboration was the result of the profession's socialization process:

> 'Within this system of stratification and domination, collaboration becomes a mechanism for both *gaining* and *sustaining* access to recognition in the professional community. Collaboration provides a means of demonstrating one's ability to those already in a position to 'recognize' others as well as keeping up one's output from such a position. Thus collaboration acts as a social regulator: it provides possible avenues of mobility for those who seek recognition; it also maintains and solidifies recognition for those who have received it.' (Beaver and Rosen, 1978: 69)

At the same time, though, professionalization entails a division of labor among researchers. This specialization augments the researcher's difficulty to address complex

scientific and technological problems individually. Hence, we hypothesize that collaborative efforts increase the reach of individual researchers, thereby allowing them to persist with their contributions. This hypothesis is now further explored in the rest of this paper.

RESEARCH SITE

We have chosen the field of transgene plants as a first case to explore the influence of relational aspects on contribution-spans. Transgene plants are a sub domain of the new biotechnology. It is one of four different biotechnology sub domains we are studying at the moment. Transgene plant research has resulted into two major application areas: (1) plant crop protection, and (2) plant quality improvement. Interest in plant quality improvement was first aroused in the 1950s as a result of the research into tissue cultures and restrictions of tissue cultures. The emergence of genetic engineering in the 1970s, combined with the specification of the Tumor Inducing Plasmid (Ti-Plasmid) in 1974, caused a renewed interest in the field. More specific, the identification of the Ti-Plasmid laid the foundations of the field that would become known as *plant genetic engineering* in the 1980s.

The first plants to be genetically engineered appeared in 1983. Ever since, transgene plant research has shown two major foci of interest. Plant crop protection aims at developing virus free plants or crops with increased stress, herbicide or disease resistance. Plant quality improvement aims at the production of hybrids and at protein improvement. Both areas have generated their first commercial products in the early 1990s. Thus, between the early 1980s and 1993, transgene plants have moved from being a scientific curiosity to a promising commercial activity.

DATA COLLECTION

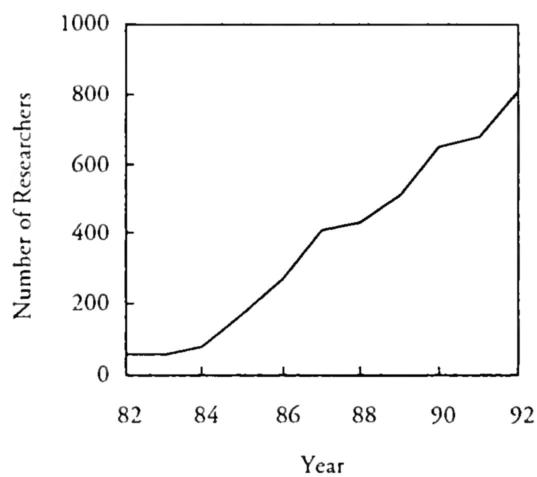
Journal articles, conference papers and patents in a given field represent a detailed, self-reported archival record of the efforts generated by researchers to solve the scientific and technological problems confronting them. Furthermore, the published literature is an appealing source of data in several respects: the publication conventions ensure a level of quality and authenticity; the data can be collected unobtrusively; the findings can be replicated and tested for reliability; and the data are publicly available and not very expensive to collect. When taken together, the literature can be viewed as a unique chronology of the efforts of researchers to establish a new field, and can provide information about the researchers involved, where they were employed, who they collaborated with, what problems they pursued, and when they were active in the field. Clearly, it would be difficult to match the comprehensive scope and longitudinal nature of the literature using other data collection techniques.

The databases of the Institute for Scientific Information (Philadelphia, U.S.) were used to identify publications related to the field of transgene plants. The databases were searched using a set of key terms on the documents that are known to be commonly

used by transgene plant researchers. The search strategy and the search results were further validated through a detailed scrutiny by three experts in the field.

The data collection procedure resulted in the identification of 1,274 unique documents related to transgene plants published between 1982 and 1992. The database was then used to identify each researcher who contributed to the field over the eleven-year period. The growth of the field, in terms of number of researchers, is shown in Figure 1. This procedure yielded a total of 2,876 researchers. A statistical database was created containing time-varying covariates for each researcher.

Figure 1
Growth of the Transgene Plant Research Community, 1982-1992



It should be noted, though, that the documents require extensive editing before they can be used as a source of data. This is necessary in order to create a consistency among author names and affiliation names. Although such a lack of standardization might not be a problem for the typical user of an electronic literature database, it would be a major source of error in determining the duration of researcher contribution-spans. Therefore, it was essential to meticulously inspect the name of each author and affiliation.

METHODS

The dependent variable for the analysis, the contribution-span, is calculated as the number of years that have elapsed from the first to the last known publication for each author. For example, if a researcher first published in 1985 and last published in 1989, the researcher's contribution span would be calculated as five years. Furthermore, it is assumed that a researcher who publishes in only one year has a contribution span of one year. As a consequence, a researcher's contribution-span is relatively unaffected by his or her frequency of publication within a given year.

Of course, calculating contribution-spans is straightforward enough. Nevertheless, certain methodological issues arise that need further explanation. It is obvious that at the moment the study is done, not all transgene plant researchers have left the field. For those researchers who are active in the field at the moment of the study, the ultimate length of their contribution-span is indeterminate. Since these individuals have not yet left the field, it is only known that the length of their contribution-span is some minimum value (i.e., the entry year to the present year). To account for this 'censoring,' event history statistics were used to analyze the data (Allison, 1990; Yamaguchi, 1991). These techniques adjust for the biases that right-censored data create.

Moreover, determining whether or not a researcher is still active in the field can be difficult in certain cases because most of them typically do not publish every year. Thus, an author's contribution-span in a field can be characterized by 'gaps' of one or more years in duration in which there are no publications to their credit. The existence of discontinuities in publication records raises the issue how frequently a researcher must publish from year-to-year in order to be considered an active contributor to the field. This issue is important to determine the proper censoring scheme to use in the analysis. The question arises: How long after someone ceases to publish is it reasonable to assume that they are no longer in the field. The answer to this question is necessary in order to determine who has exited the field and who continues to be a participant.

Inspection of the dataset showed that less than five percent of the researchers present have a gap between publications of longer than three years in duration, and still fewer have an exceptionally large gap such as five years or more. These large gaps may be indicative of an individual who does not contribute continuously to the field. Thus, although his/her contribution-span would seem long, it is not indicative of his/her actual participation in the field. Therefore, we considered all contribution-spans with gaps of no more than three years as contiguous. If a researcher had a gap longer than three years, we treated it as if the case exited and then subsequently re-entered the field.

Having determined the contribution-spans, it is interesting to examine the factors affecting them. To this end, different analytical approaches are possible. Using the literature, a number of explanatory variables were constructed. As outlined in the previous sections of this paper, we were primarily interested in constructing covariates that capture the relational dimensions in the researcher community under investigation. These covariates were computed using the sociometric network analysis package STRUCTURE™ (Version 4.2) developed by Burt at Columbia University (Burt, 1991). In the next section, we provide a detailed overview of the different relational covariates used in the present analyses.

A first approach to implement the event history analysis is to use a 'single-spell' model. In this approach, the value for each explanatory variable is taken according to the last year in the author's contribution-span. The 'single-spell' approach was implemented using the LIFEREG procedure of SAS™ and using the LIMDEP™ software package (version 6.0) developed by Greene at New York University. This approach was chosen for all 2,876 unique authors in the original database. The results of this 'single-spell' approach for the complete dataset are reported below.

In addition, the explanatory variables can be treated as time-varying covariates. In this second approach, each explanatory variable has values that vary in the course of an author's contribution-span. From a computational perspective, though, this 'multiple-spell' approach is much more complex. The LIMDEP™ program allowed us to examine the factors affecting researcher contribution-spans (with time-varying co-variates) for 689 unique researchers in the database. Hence, the results of this 'multiple-spell' approach are also reported below, only now for a sample of 689 researchers from the original database (note: T-test and WILCOXON non-parametric comparisons showed that the distributions of contribution-spans of the 689 researcher sample did not differ significantly from the contribution-span distribution in the complete 2,876 researcher population). It should also be noted that we are currently installing a more powerful event history package on a mainframe computer (RATE™) which will allow us to run the 'multiple-spell' approach on the complete 2,876 author dataset.

Besides the sociometric covariates, only a few other explanatory variables were introduced in the present analyses. First, the kind of organization in which each author is employed was coded according to whether they reside in an academic, government, established industrial, or new biotechnology company research laboratory. Given the tremendous growth in the number of new biotechnology firms over the last years, we found it necessary and helpful to distinguish between researchers residing in the research laboratories of (mostly large) established firms (e.g. MONSANTO) and their colleagues employed by those new start-ups (e.g. Plant Genetic Systems). These organization variables were introduced as dummy variables in the explanatory analyses.

Second, at the individual-level, a variable was constructed to reflect an author's cumulative productivity in the field measured by the cumulative number of publications to his/her credit.

Third, at the community-level (in order to capture the collective aspects of the knowledge creation and diffusion processes in the researcher community) two variables were created. One reflects the population size of the transgene plants field in each year. Population size is measured in terms of the number of individual authors who publish in the field in a given year. In addition, a second-order variable, the square of population size, was introduced in order to capture any quadratic association between population size and contribution-span.

The other covariates were then constructed to capture the relational dimensions of the knowledge creation and diffusion process.

INDICATORS OF RELATIONAL DIMENSIONS OF KNOWLEDGE CREATION

As stated previously, we are mainly interested in investigating the relationship between relational covariates and author contribution-spans. Relational covariates can be determined by analyzing the authors' sociometric position in the network of co-authorships that emerges from the original database. In order to define and compute these sociometric indicators, we used the theory and the programs developed by Burt (1976, 1980 & 1992) at Columbia University. Although an overwhelming number of

sociometric indicators has been developed over the last decade (e.g. Freeman, 1977&1979; Knoke and Kuklinski, 1990; Rogers and Kincaid, 1981; Yamagishi et al., 1988), the indicators used in the present analysis are based on Burt's approach to sociometric modeling.

The network variables were computed using the JEDIT module of STRUCTURE™ (Burt, 1991). The JEDIT module allows one to analyze joint involvement data. It infers relations from involvement in the same events, or affiliations with the same groups. In this approach, two actors i and j are tied together to the extent that they are involved or affiliated with the same events or groups. Where scientific articles are events and the authors of the article are actors, z_{ij} measures the connection between persons i and j in the metric of number of co-authored articles. These connections are subsequently used to compute different sociometric indices. For the analyses reported in this paper, the sociometric indices were treated as time-varying covariates. This means they were computed for each year (or number of years) of an author's contribution-span.

The time-periods used to compute the time-varying network variables were, in part, determined by the computational limitations of the program (which pose constraints on the maximum number of actors in the network to be analyzed simultaneously). Hence, the network indices were computed for the periods 1982-84, 1985-87, 1988-89, and the years 1990, 1991, and 1992. Thus, an author who is active in the years 1982, 1983 and 1984 will have the same value for the sociometric index in each year. If the author is still active in 1985, his or her sociometric indices may change, though. This aggregation of years into longer time-periods is, of course, to a certain extent arbitrary. However, sensitivity analyses on the time periods used did not reveal any major alterations in the results discussed below.

First, five variables (called autonomy variables) are produced to describe author i's network, analyzed as ego. These variables provide measures of the size, the density, and the brokerage opportunities in an individual's network.

The first variable simply reflects the ego-network, N. It is a count of ego's contacts [CTACTS]. This is everyone who is connected with i. It is the size of i's network. In the present analyses, this variable is equal to the number of co-authors each individual author has during the time-periods considered in the analyses.

The second variable, nonredundant contacts, is a count of the number of independent contacts in ego's network [NONRCTACTS]. The meaning of this variable is easy to grasp. Let us assume that author i appears on one particular article with two co-authors, j and k. In this case, the first contact variable [CTACTS] would equal 2. It is simply the number of co-authors with whom i is involved. The second variable [NONRCTACTS], though, will equal 1 for author i, since both j and k are strongly tied. Thus, by counting the connection between i and j, the connection between i and k is implicit given the strong connection between j and k. The distinction between both variables describing the author's network is thus based on the concept of transitive relationships.

The third variable describing ego's network is called contact efficiency [CTACTEFF]. It is the ratio of the nonredundant contacts to all contacts involving author i.

The fourth variable describes the network density [NDENS]. It is the average marginal strength of relations between contacts (Burt, 1991):

$$NDENS=[\Sigma_j \Sigma_q z_{jq}/\max(z_{jk})]/[N(N-1)], \qquad j \neq q$$

where $\max(z_{jk})$ is the largest of j's relations to anyone, so density ranges from 0 (no relations between contacts) to 1 (maximum strength relations between all contacts).

The fifth variable is a proportional density measure [PRDENS]. It is the proportion of contact pairs that have some kind of connection with one another (Burt, 1991):

$$PRDENS=\Sigma_j \Sigma_q \partial_{jq}/N(N-1), \quad j \neq q$$

where ∂_{jq} is 1 if z_{jq} is nonzero, otherwise ∂_{jq} equals 0. This measure of density varies from 0 (no relations between contacts) to 1 (every pair of contacts is connected). In a network of contacts all connected by weak relations, NDENS is low and PRDENS is high. Given the nature of our data (strong ties based on co-authorships), both density measures can be expected to be correlated.

For the co-authorship data used in the present analyses, the correlations between the five network indices are shown in Table 1.

Table 1
Correlations among Five Indicators of Ego's Autonomy Position in the Transgene Plant Co-Author Network

	CTACTS	NONRCTACTS	CTACTEFF	NDENS	PRDENS
CTACTS	1.00**	0.62**	-0.25**	-0.37**	-0.29**
NONRCTACTS		1.00**	0.24**	-0.61**	-0.51**
CTACTEFF			1.00**	-0.08**	-0.07**
NDENS				1.00**	0.89**
PRDENS					1.00**

*: $p<0.01$; **: $p<0.001$

Whereas the previous sociometric indices concern the freedom or autonomy a researcher has to act within the limited sphere of his/her contacts, power or prominence concerns the author's ability to dominate the community. Four different indices were computed reflecting a researcher's power position in the co-author network. The first index is the author's choice status [STATUS]. It is the number of co-authors to whom an author is connected divided by the number to whom he or she could have been connected (Burt, 1991):

$$\text{STATUS of } i=\sum_j \partial_{ij}/(N-1), \qquad j\neq i$$

where N is the number of authors in the whole community (not just those connected to i), and ∂_{ij} equals 1 if j can reach i ($z_{ij}>0$), otherwise ∂_{ij} equals 0. STATUS varies from 0 (when no one reaches i) to 1 (when everyone else reaches i).

This can be misleading when two individuals are the object of relations from the same number of others, but one receives weak relations while the other receives strong relations. Therefore, Burt's second power index weights the relationships (read: co-authorships) by their strength (where max (z_{jk}) is j's strongest relation to anyone else):

$$\text{EXTREL=extensive relations to } i= \sum_j[z_{ji}/\max(z_{jk})]/(N-1), \qquad j\neq i,k$$

This variable ranges from 0 (when i receives no relations) to 1 (when i is the object of a maximum strength relation from every actor). Given the nature of our data structure (strong ties based on co-authorships), we can expect both power indicators to be strongly correlated (see also Table 2). Therefore, only one of them is used in the event-history analyses.

Two additional power indices are computed by STRUCTURE™. One of them is the author's exclusive prominence in the community [EXCLREL]. This index is relevant because a member of a highly connected clique has a high score on extensive prominence [EXTREL], but so does everyone else in the clique. Therefore, the exclusive prominence indicator measures the extent to which author i is the object of exclusive relations from everyone. It varies from 0 (when i receives no relations) to 1 (when i is the only contact for every other actor). It distinguishes between the 'top' and the 'bottom' of the social structure in a given system.

Finally, we computed the 'ultimate' power of each author [POWER] as the extent to which he/she is the object of exclusive relations from powerful others. The correlations between the four power variables are shown in Table 2.

Table 2
Correlations among Four Indicators of Ego's Power Position in the Transgene Plant Co-Author Network

	STATUS	EXTREL	EXCLREL	POWER
STATUS	1.00**	0.97**	0.70**	0.31**
EXTREL		1.00**	0.75**	0.39**
EXCLREL			1.00**	0.46**
POWER				1.00**

*: $p<0.01$; **: $p<0.001$

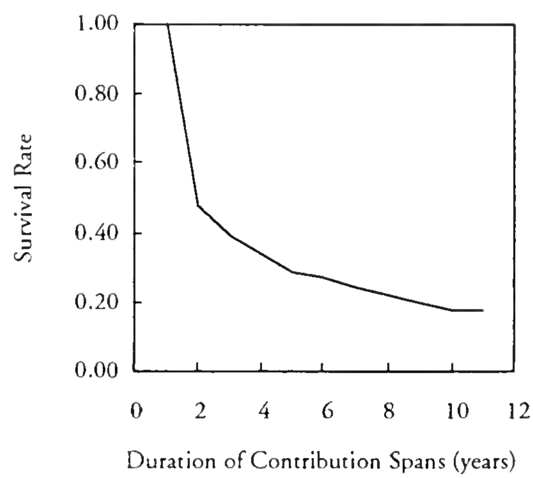
RESULTS

Non-parametric estimates of the survival and hazard functions

Using the published literature data on transgene plants for the period 1982-1992, the contribution-span for 2,876 researchers and several explanatory variables associated with each author were compiled into a statistical database. The non-parametric survival and hazard rates were analyzed using the LIFETEST procedure of SAS. Of the 2,876 cases, 1,303 (45.3%) were active within the last two years of the data and were classified as censored.

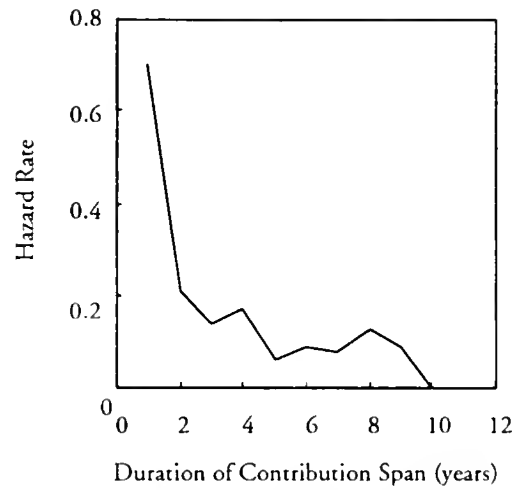
Using the LIFETEST procedure, the first step in the analysis consisted of making non-parametric estimates of the survival and hazard functions for the data. The lifetable approach was chosen. The results of this procedure are shown in Figures 2 and 3. The median survival time is about two years: that is, half the sample has left the field within two years of their first publication. The survival time diminishes rapidly and levels off at about 0.20 for contribution-spans of eight years or more.

Figure 2
Non-Parametric Estimate of Survival Function for Researcher Contribution Spans in Transgene Plants



The hazard function is also negatively-sloped (see Figure 3). The hazard rate decreases very rapidly for researchers who have contribution-spans of at least two years: that is, the probability of a researcher ceasing to contribute after having contributed for two years is only about 0.20, compared to about 0.70 for a researcher in the field for only 1 year. The basic implication of the hazard function is that the longer a researcher contributes to the field, the less likely it is that he or she will leave the field. As is obvious from Figure 3, the risk of leaving the field is highest within the first year.

Figure 3
Non-Parametric Estimate of the Hazard Function for Researcher Contribution Spans in Transgene Plants



Parametric models of contribution-spans: the single-spell approach

The next step in the analysis was to determine the parametric model that best fits the distribution of contribution-spans. The basic model adopted for this analysis was:

$$Y = X\beta + \sigma\varepsilon$$

where Y is the log of the contribution-span, X is the matrix of explanatory variables, β is a vector of unknown regression parameters, σ is a scale parameter and ε is a vector of errors from an assumed distribution. Specifically, we evaluated three different types of distributions: the exponential, Weibull, and log-logistic distribution. The parameters were estimated by maximum-likelihood using a Newton-Raphson algorithm. The overall fit of a model is represented by the log-likelihood function. On the basis of the log-likelihood score, the log-logistic distribution appeared to provide the best overall fit. Hence, this distribution was chosen as the basis for estimating the regression coefficients of the explanatory variables in the model (see Table 3).

The model was estimated both with the LIFEREG procedure of SAS and using LIMDEP. The results of these analyses are shown in Table 3. Among the dummy variables that control for organization type, the variables that distinguish academic laboratories and the laboratories of established industrial firms from other types of organizations are both significant ($p<0.05$). The positive coefficients suggest that researchers employed in academic or established firm laboratories have longer contribution-spans compared to researchers with other types of affiliations. More specifically, the new technology based firm dummy variable did not attain statistical

significance. Although these types of organizations are often claimed to play a vital role in the development of a new technology, especially in biotechnology, researchers employed at these firms do not appear to have longer contribution-spans than their colleagues at other organizations.

At the community-level, we find the first- and second-order population terms to be statistically significant ($p<0.001$). The negative coefficient for population size combined with the positive coefficient for the second-order term implies a U-shaped relationship between population size and researcher contribution-spans. This result supports previous research which suggests that *'there may be a point of critical mass for the community, where its size is sufficiently large to become significant and thereby increase contribution-spans'* (Rappa and Garud, 1992).

At the individual-level, an author's cumulative productivity is highly significant ($p<0.001$). Thus, and as might be expected, researchers who accumulate a greater number of publications will have longer contribution-spans.

Perhaps more interesting are the relational variables that have been included in the analyses. Three out of the five 'autonomy' indices attain statistical significance (NONRCTACTS, $p<0.01$; CTACTEFF, $p<0.05$, NDENS, $p<0.05$). They indicate that an individual author's position in a network of co-authors influences his or her contribution-span. More specifically, the more a researcher is embedded in communal collaborations with his peers in the community, the higher his longevity in the field. Also, network density is positively related to researcher contribution-spans.

The importance of network positions on contribution-spans is further emphasized by the 'power' indices. Exclusive prominence (EXCLREL) is positively related to a researcher's contribution span ($p<0.001$) suggesting that a researcher's contribution-span increases when he belongs to the 'top' of the social structure within his or her field. The only variable which does not behave as one might expect is the STATUS-variable. Although it is statistically significant ($p<0.01$), its negative coefficient requires further exploration.

Parametric models of contribution-spans: the multiple-spell approach

Introducing time-varying covariates is the next logical step in the present analysis. In a multiple-spell approach, the covariates are allowed to take on different values for each year the author is active in the field. Due to their computational complexity, in particular when the dataset is fairly large, multiple-spell models are only rarely used. For instance, for the 2,876 researchers in the database, a multiple-spell model results in 4,833 spells.

We implemented the multiple-spell approach using LIMDEP. Due to computational limitations, both the number of covariates and the number of spells had to be restricted. The results of this exploratory analysis are reported in Table 4.

Although the multiple-spell results should be considered preliminary (only 689 unique authors out of the 2,876 author-database could be used, resulting in 1,104 spells), it is interesting to note that the findings reported in Table 4 are in line with the ones reported in the single-spell model of Table 3.

Table 3
ML Estimation of Contribution-Spans using a Single-Spell Approach

VARIABLES	LOG-LOGISTIC DISTRIBUTION
Type of organization	
ACADEMIA	0.044**
	(0.016)
ESTABLISHED FIRM	0.058*
	(0.025)
NEW BIOTECHNOLOGY FIRM	0.006
	(0.028)
Community-level	
POPULATION	-0.002***
	(0.000)
$\text{POPULATION}^2/1000$	0.003***
	(0.000)
Individual-level	
CUM. PRODUCTIVITY	0.524***
	(0.005)
Relational	
CTACTS	-0.006
	(0.007)
NONRCTACTS	0.039**
	(0.013)
CTACTEFF	0.117*
	(0.048)
NDENS	0.193*
	(0.103)
PRDENS	0.041
	(0.093)
STATUS	-15.926**
	(5.017)
EXCLREL	0.301***
	(0.090)
POWER	-0.005
	(0.003)
Scale parameter	0.160
	(0.002)
LOG-LIKELIHOOD	-829

Notes:
1. Total number of researchers=2,867 (1,303 right-censored)
2. Significance level: *: $p<0.05$ — **: $p<0.01$ — ***: $p<0.001$
3. Only the 'best-fitting' distribution is shown (other distributions tested: Weibull, Exponential)
4. Figures in parentheses are standard errors of estimates

Table 4
ML Estimation of Contribution-Spans using a Multiple-Spell Approach

VARIABLES		WEIBULL DISTRIBUTION
Type of organization		
	ACADEMIA	0.223***
		(0.056)
	ESTABLISHED FIRM	0.271**
		(0.084)
	NEW BIOTECHNOLOGY FIRM	0.193
		(0.108)
Community-level		
	POPULATION	0.000
		(0.000)
	POPULATION^2/1000	0.001**
		(0.000)
Individual-level		
	CUM. PRODUCTIVITY	0.406***
		(0.032)
Relational		
	NONRCTACTS	0.048*
		(0.030)
	NDENS	-0.468***
		(0.109)
	STATUS	-5.153
		(8.306)
	EXCLREL	0.646*
		(0.307)
Scale parameter		0.459
		(0.026)
LOG-LIKELIHOOD		-580

Notes:

1. Total number of spells= 1,104 (689 unique researchers, 305 right-censored)
2. Significance level: *: $p<0.05$ — **: $p<0.01$ — ***: $p<0.001$
3. LIMDEP only allows Weibull and Exponential distributions for this approach. Weibull provided the best fitting distribution hence, these results are presented
4. Figures in parentheses are standard errors of estimates

As far as organization-type is concerned, authors in academic laboratories or researchers residing in established firms contribute longer than their peers in other research organizations ($p<0.01$). The author's cumulative productivity has the same significant effect as in the single-spell model ($p<0.001$). The relational variables included in the analysis are all statistically significant ($p<0.05$), except for the STATUS variable. The directions of their influence are to a large extent similar to the ones reported with the single-spell model (see Table 3), except for the network density variable which has a negative coefficient. This is due to the multiple-spell approach. Indeed, the longer the contribution-span of the author (over the period 1982-1992), the larger the size of the community has grown (see Figure 1) and, almost naturally, the smaller the network density. The multiple-spell approach unveils this dynamic, whereas the single-spell approach does not. Also, the result for the first-order term of the population variables is slightly different from the ones reported in Table 3.

To conclude, although the multiple-spell results are promising, we should keep in mind that, at present, they were obtained from a limited sample of the total population with only a restricted number of explanatory, albeit time-varying, covariates included in the model.

Discussion and Conclusion

In this paper, we have investigated a number of determinants of contribution-spans of researchers in the field of transgene plants, using the published literature as a source of data. Non-parametric estimates of the survival rate and hazard rate were made, and it was found that a majority of researchers tend to leave the field within two years of their first publication, while those who persist are less likely to leave the field at all. The fact that it is the early years that are most critical for a researcher to leave a field is understandable, since at that point not much of a researcher's career has been invested in the field.

An examination of the relationship between several covariates and the length of contribution-spans indicates the importance of organization-type: researchers residing in academic laboratories or in the laboratories of (mostly large) established firms tend to contribute longer than, say for instance, their colleagues employed by new biotechnology firms. Not astonishingly, an author's cumulative productivity has a positive effect on his contribution-span. At the community-level, population size appears to have effects that should not be neglected.

The major focus of the paper, though, was to start investigating the relationship between the relational dimensions of a research community and researcher contribution-spans. To this end, sociometric network techniques were used to compute a range of variables reflecting the authors' embeddedness within their research community. Co-authorship data provided the starting-point to compute sociometric indicators. Notwithstanding the preliminary nature of the results reported in the paper, the findings lend support to the hypothesis that a researcher's position in the community's co-author network is an important determinant of his/her contribution-

span. Thus, the better a researcher is embedded in research collaborations within a field, the higher the likelihood he will persist.

The analyses were pursued using single- and multiple-spell approaches. It is obvious that the multiple-spell approach is the methodologically more correct one. For the present analyses, both approaches yielded comparable results. However, the multiple-spell approach is currently being explored in greater detail.

Finally, we believe the application of sociometric and event history techniques to bibliometric data offers new perspectives to the management of R&D activities. Indeed, analyzing contribution-spans may eventually serve as a useful approach to unveil the fundamental dynamics of new technology development. By focusing on the determinants of researcher contribution-spans, our aim is to shift attention away from predicting the technological future and towards understanding the fundamentals of researcher behavior. Improvements in our understanding of survival and hazard rates for researchers in a field may ultimately lead to the identification of critical factors and events (e.g. with regard to networking, cfr. the E.C. Human Capital and Mobility Program) that can inform our decisions regarding emerging technologies. A comprehensive and systematic insight into the sustained commitment of researchers to the ideas they are pursuing may therefore prove invaluable.

References

Allen, T.J. 1966. 'Studies of the problem-solving process in engineering design,' *IEEE Transactions on Engineering Management*, Vol. 18, pp. 72-83.

Allison, P.D. 1990. Event History Analysis. Applied Social Research Methods Series, Vol. 46, Newbury Park: Sage Publications (fifth printing).

Beaver, D. and R. Rosen. 1978. 'Studies in scientific collaboration,' *Scientometrics*, Vol. 1, pp. 65-84.

Ben-David, J. and R. Collins. 1966. 'Social factors in the origins of a new science: the case of psychology,' *American Sociological Review*, Vol. 31, pp. 451-465.

Burt, R.S. 1976. 'Positions in networks,' *Social Forces*, Vol. 55, pp. 93-122.

Burt, R.S. 1980. 'Models of network structure,' *Annual Review of Sociology*, Vol. 6, pp. 79-141.

Burt, R.S. 1991. *Structure Reference Manual: Version 4.2.* Columbia University's Center for the Social Sciences.

Burt, R.S. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge, Mass.: Harvard University Press.

Collins, H.M. 1974. 'The TEA-set: tacit knowledge and scientific networks,' *Science Studies*, Vol. 4, pp. 165-186.

Constant, E.W. 1980. *The Origins of the Turbojet.* Baltimore: The Johns Hopkins University Press.

Crane, D. 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities.* Chicago: The University of Chicago Press.

Debackere, K. and M.A. Rappa. 1993. 'Institutional variations in problem-choice and persistence among researchers in an emerging field,' forthcoming in *Research Policy*.

Dosi, G. 1982. 'Technological paradigms and technological trajectories,' *Research Policy*, Vol. 11, pp. 147-162.

Edge, D.O. and M.J. Mulkay. 1976. *Astronomy Transformed: The Emergence of Radio Astronomy in Britain*. New York: John Wiley & Sons.

Freeman, L.C. 1977. 'A set of measures of centrality based on betweenness,' *Sociometry*, Vol. 40, pp. 35-41.

Freeman, L.C. 1979. 'Centrality in social networks. I: Conceptual clarifications,' *Social Networks*, Vol. 1, pp. 215-239.

Gieryn, T.F. 1978. 'Problem retention and problem change in science,' *Sociological Inquiry*, Vol. 48, pp. 96-115.

Gieryn, T.F. and R.F. Hirsch. 1983. 'Marginality and innovation in science,' *Social Studies of Science*, Vol. 13, pp. 87-106.

Granovetter, M.S. 1973. 'The strength of weak ties,' *American Journal of Sociology*, Vol. 78, pp. 1360-1380.

Hagstrom, W.O. 1965. *The Scientific Community*. New York: Basic Books Inc.

Hughes, T.P. 1989. *American Genesis: A Century of Invention and Technological Enthusiasm*. New York: Viking.

Katz, R. and T.J. Allen. 1982. 'Investigating the Not-Invented-Here (NIH) syndrome: a look at the performance, tenure and communication patterns of 50 R&D groups,' *R&D Management*, Vol. 12, pp. 7-19.

Knoke, D. and J.H. Kuklinski. 1990. *Network Analysis*. Applied Social Research Methods Series, Vol. 28, Newbury Park: Sage Publications.

Kuhn, T.S. 1970. *The Structure of Scientific Revolution*. Chicago: The University of Chicago Press (2nd ed.).

Laudan, R. 1984. *The Nature of Technological Knowledge*. Dordrecht: Reidel.

Layton, E. 1974. 'Technology as knowledge,' *Technology and Culture*, Vol. 15, pp. 31-41.

Layton, E. 1977. 'Conditions of technological development,' in *Science, Technology and Society: A Cross-Disciplinary Perspective*, I. Spiegel-Rösing and D. de Solla Price (eds.), London and Beverly Hills: Sage Publications.

Nelson, R.R. and S.G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Harvard University Press.

Pelz, D.C. and F.M. Andrews. 1966. *Scientists in Organizations: Productive Climates for R&D*. New York: John Wiley.

Polanyi, M. 1958. *Personal Knowledge*. London: Routledge and Kegan Paul.

Rappa, M.A., Debackere, K. and R. Garud. 1992. 'Technological progress and the duration of contribution spans,' *Technological Forecasting and Social Change*, Vol. 42, pp. 133-145.

Rappa, M.A. and K. Debackere. 1992a. 'Monitoring progress in R&D communities,' in *Representations of Science and Technology*, P. Weingart, R. Sehringer en M.

Winterhager (eds.). Leiden: DSWO Press, Lisbon Institute, University of Leiden, pp. 253-265.

Rappa, M.A. and K. Debackere. 1992b. 'Technological Communities and the Diffusion of Knowledge,' *R&D Management*, Vol. 22, No. 3, pp. 209-220.

Rappa, M.A. and R. Garud. 1992. 'Modeling contribution-spans of scientists in a field: the case of cochlear implants,' *R&D Management*, Vol. 22, No. 4, pp. 337-348.

Rogers, E.M. and D.L. Kincaid. 1981. *Communication Networks: Toward a New Paradigm for Research*. New York: MacMillan.

Root-Bernstein, R.S. 1989. *Discovering: Inventing and Solving Problems at the Frontier of Scientific Knowledge*. Cambridge, Mass.: Harvard University Press.

Rosenberg, N. 1982. *Inside the Black Box*. Cambridge, UK: Cambridge University Press.

Sahal, D. 1981. *Patterns of Technological Innovation*. Reading, Mass.: Addison-Wesley Publishing Co.

Small, H. and E. Greenlee. 1986. 'Collagen research in the 1970s,' *Scientometrics*, Vol. 10, pp. 95-117.

Stewart, J.A. 1990. *Drifting Continents and Colliding Paradigms*. Bloomington: Indiana University Press.

Tushman, M.L. and P. Anderson. 1986. 'Technological discontinuities and organizational environments,' *Administrative Science Quarterly*, Vol. 31, pp. 439-465.

Yamagishi, T., Gillmore, M.R. and K.S. Cook. 1988. 'Network connections and the distribution of power in exchange networks,' *American Journal of Sociology*, Vol. 93, pp. 833-851.

Yamaguchi, K. 1991. *Event History Analysis*. Applied Social Research Methods Series, Vol. 28, Newbury Park: Sage Publications.

Ziman, J. 1987. *Knowing Everything about Nothing*. Cambridge, UK: Cambridge University Press.